

Eine Dokumentation maschinenlesbarer Textkorpora der deutschen Gegenwartssprache

0. Vorbemerkung

Die steigende Zahl empirisch-deskriptiv orientierter linguistischer Arbeiten, unter denen das "Projekt eines großen interdisziplinären Wörterbuchs der deutschen Sprache" wiederum einen hervorragenden Platz einnimmt¹, das wachsende Interesse an einem wissenschaftlich begründeten Aufbau sowie an einer konsistenten Methodik der Auswertung von Textkorpora auf der einen Seite und die ungenügende Kenntnis von Umfang, Inhalt und Gebrauchswert vorhandener maschinenlesbarer Textkorpora auf der anderen Seite waren der Anlaß zu der hier vorgelegten Dokumentation.

Ich habe mich bei der Auswahl auf jene Textkorpora beschränkt, die Repräsentationen deutscher Gegenwartssprache darstellen, für die maschinelle Auswertung eingerichtet worden sind und die mindestens 200.000 Wortstellen umfassen, wobei diese Zahl keine Aussage über einen notwendigen Mindestumfang von Textkorpora bedeuten soll.²

Die Angaben für die hier vorgelegte vergleichende Übersicht habe ich den in Kapitel 4 aufgeführten Publikationen entnommen. Dort nicht enthaltene Auskünfte verdanke ich den für die Konzeption und Betreuung der einzelnen Korpora verantwortlichen Kollegen.

Die hier vorgelegte Dokumentation stellt eine überarbeitete, um die englischsprachigen Korpora gekürzte, gleichzeitig um einige deutschsprachige Korpora erweiterte und in etlichen Punkten aktualisierte Neufassung meines Beitrags Schaefer (1976) dar.

1. Über den wissenschaftlichen Gebrauchswert maschinenlesbarer Textkorpora

Unter einem Textkorpus verstehen wir jede endliche Menge von natürlich-sprachlichen Texten, die für die Zwecke sprach- oder literaturwissenschaftlicher Untersuchungen zu einer Sammlung zusammengestellt werden.³ Dabei bestimmen das Vorverständnis vom Objektbereich und das Erkenntnisinteresse, ob die Texte einer solchen Kollektion

- systematisch zusammengestellt und geordnet sind,
- einer oder mehreren Sprachen entstammen,

- nach diachronischen oder synchronischen Gesichtspunkten ausgewählt werden,
- außer der Standardsprache auch diatopische und/oder diastratische Varietäten repräsentieren,
- Manifestationen der geschriebenen und/oder gesprochenen Sprache darstellen.

Die Frage des Werts oder Unwerts eines Textkorpus für eine linguistische Theorienbildung läßt sich nicht grundsätzlich, sondern immer nur von bestimmten theoretischen Positionen aus beantworten.

Aufgrund der sie jeweils konstituierenden Methoden wollen wir eine axiomatisch-deduktive und eine empirisch-induktive bzw. deskriptive Linguistik unterscheiden, wobei wir gleichzeitig zugestehen, daß diese Dichotomie insofern hypothetischen Charakter hat, als beide Typen innerhalb unserer Wissenschaft nicht in reiner Form, sondern immer nur als primär-deduktive und primär-induktive Linguistik repräsentiert sind. Daß beide Theorien auf Verfahrensschritte der jeweils anderen Theorie rekurrieren, hat seinen Grund im Objektbereich Sprache, der zum einen immer schon empirisch gegeben ist, über den zum anderen über eine bloße Aufzählung von Fakten hinaus Aussagen allgemeingültigen Charakters formuliert werden sollen.⁴

Textkorpora dienen in der empirisch-deskriptiven Linguistik als unerläßliche Datenbasis für die Formulierung von Hypothesen über die Struktur des Objektbereichs sowie als Material für die Beschreibungs- und grammatischen Argumentationsverfahren und für die jeweiligen Bestätigungsprozeduren.

Gegen die Verwendung solcher notwendigerweise immer limitierten Textkorpora wenden sich die Vertreter einer axiomatisch-deduktiven Linguistik, die "nicht gegebene Sprech- oder Schreibereignisse, sondern die intuitiven Auffassungen der Sprecher über die Form von grammatisch richtigen Sätzen, die solchen Ereignissen zugrunde liegen", beschreiben möchten.⁵

Dieses Ziel zu erreichen, leisten Textkorpora ihrer Meinung nach keine Hilfe: "Eine Aufzeichnung natürlicher Rede zeigt stets zahlreiche falsche Ansätze, Abweichungen von Regeln, Abänderungen der Strategie mitten im Sprechen usw. [...]. Beobachtungen des Sprachgebrauchs oder Hypothesen über die Anlagen, sprachlich zu reagieren, über Gewohnheiten usw. können wohl Evidenzen für die Beschaffenheit dieser mentalen Realität beibringen, sie können aber sicherlich nicht den tatsächlichen Gegenstand der Linguistik ausmachen"⁶, den in der Theorie einer axiomatisch-deduktiven Linguistik die Grammatik als die "Beschreibung der immanenten Sprachkompetenz des idealen Sprecher-Hörers" darstellt.⁷

Auch wenn die von den Vertretern einer axiomatisch-deduktiven Linguistik innerhalb ihrer Argumentationen aufgeführten Beispiele des Sprachgebrauchs eher einen die Theorie illustrierenden Charakter haben und nicht das Material für Hypothesen oder Bestätigungen über die Struktur des zu untersuchenden Gegenstandes darstellen, bleibt für jede axiomatisch-deduktive Linguistik die Frage zu beantworten, "wie die Beobachtungsdaten, die durch die Grammatik erklärt werden sollen, zu gewinnen sind"⁸.

Denn abgesehen davon, daß die Entwicklung einer solchen Grammatik nicht auf jene Empirie verzichten kann, die bereits die Formulierung der Prämissen anleitet, muß eine axiomatisch-deduktive Linguistik die Adäquatheit ihrer Konklusionen in Bezug auf den Objektbereich Sprache mithilfe entsprechender Kongruenzregeln überprüfen. Auch dazu benötigt sie umfangreiche Kenntnisse über die vielfältigen Formen des Sprachgebrauchs, die in jedem Fall die Empirie liefert.

Wenn als empirisches Material Textsammlungen gleich welchen Umfangs und gleich welcher Auswahl abgelehnt werden, weil sie einmal "unkontrolliert nebeneinander ganz unterschiedliche Erscheinungen enthalten: grammatische und ungrammatische Sätze, grammatische Abweichungen der verschiedensten Art, die zu einer ganz wertlosen Grammatik führen würden, wenn man sie alle berücksichtigte"; da Korpora zum anderen, "streng genommen, keinerlei Feststellung über Grammatikalität, strukturelle Ähnlichkeit, Ambiguität usw." enthalten⁹, bleibt schließlich allein der Rückgriff auf Ideokompetenz bzw. auf das mit ihrer Hilfe jeweils produzierte Korpus, dem auf vielleicht etwas gehobenerem Niveau alle jene Mängel anhaften, die nach Meinung der Vertreter einer axiomatisch-deduktiven Linguistik Textsammlungen nun einmal eigen sind.

Wie wir schon zu Beginn dieses Kapitels ausgeführt haben, läßt sich die Frage nach dem wissenschaftlichen Gebrauchswert eines Textkorpus nicht grundsätzlich, sondern immer nur im Rahmen einer bestimmten Theorie beantworten. Eine empirisch-deskriptive bzw. primär-induktive Linguistik, deren vornehmliches Ziel die Beschreibung des Sprachgebrauchs ist, kann auf die Verwendung eines in Umfang und Auswahl dem Objektbereich angemessenen Textkorpus nicht verzichten. Da auch die Beschreibung der Sprachfähigkeit, will sie die Forderung der Adäquatheit erfüllen, nicht ohne die Beobachtung der vielfältigen Formen des Sprachgebrauchs auskommen kann, sollten auch die Vertreter einer primär-deduktiven Linguistik die Verwendung eines Textkorpus nicht so strikt, wie dies bisweilen geschieht, ablehnen. "For example, one whereby the introspections of the linguist, as his own informant, are accorded a provisional measure of evidential validity, but are held answerable ultimately to more objective confirmation, provided by corpus evidence and informant tests."¹⁰

Insgesamt eignet sich ein Textkorpus, das in Umfang und Auswahl im Hinblick auf seine intendierte Verwendung bestimmten Kriterien der Repräsentativität genügt,

1. als Material zur Formulierung von Hypothesen bzw. Prämissen über den Objektbereich;
2. als unerläßliche Datenbasis für die grammatische und lexikalische Beschreibung als einer komplexen Handlung, die sich aus den Teilaktivitäten des Beobachtens, des innersprachlichen Vergleichens und des Identifizierens, der Delimitation bzw. des Segmentierens und schließlich des Klassifizierens besteht¹¹;
3. als Material für das Verfahren des grammatischen Argumentierens;
4. als Material für die Bestätigung bzw. das Verwerfen von Hypothesen sowie zur Überprüfung der Adäquatheit der aus bestimmten Prämissen gefolgerten Konklusionen;
5. als Sammlung von Belegen zur Illustration bestimmter Gebrauchsweisen von Wörtern und Syntagmen.

Im Unterschied zu einer Belegsammlung, deren Zusammenstellung in der Regel durch das Vorverständnis des Bearbeiters vom Objektbereich bestimmt und häufig am eingeschränkten Zweck einer vorweg festgelegten Verwendung orientiert ist, bietet ein Textkorpus ein in Umfang und Gestalt nicht auf bestimmte eingegrenzte Problemstellung zugeschnittenes, vom Vorverständnis einzelner Bearbeiter unabhängiges, die Ideokompetenz stützendes, ergänzendes bzw. kontrollierendes Material.

Textkorpora, die wie die hier vorgestellten zusätzlich auf maschinenlesbaren Datenträgern verfügbar sind, beanspruchen für ihre Aufbewahrung nicht nur einen vergleichsweise geringen Platz, sie ermöglichen vor allem einen schnellen Zugriff zum gesuchten Material, darüber hinaus ihre beliebig häufige Vervielfältigung und eine Reihe per Hand kaum zu leistender Sortierungen, wie sie z.B. alphabetisch vor- und rückläufig sortierte Wortformenlisten, Ranglisten und verschiedene Arten von Indices und Konkordanzen darstellen. Insgesamt bilden solche Sortierungen rudimentäre Formen sogenannter Korpuswörterbücher, die als Computerausdrücke oder auf Mikrofilm oder -fiche aufgenommen ein überaus hilfreiches Material für grammatische und lexikalische Arbeiten zur Verfügung stellen.

2. Kriterien der Auswahl von Texten für ein Korpus

Ein zentrales Problem jeder auf eine Korpusauswertung angewiesenen Linguistik ist der Grad der Repräsentativität der ein Korpus bildenden Auswahl aus der Gesamtheit sprachlich realisierter Äußerungen. Da die Erfassung und erst recht die Auswertung aller auch in einem sehr be-

grenzten Zeitraum produzierten Sprachverwendungen aller Mitglieder auch nur einer Sprachgruppe unmöglich zu leisten ist, wird man für die Erstellung eines selbst sehr umfangreichen Korpus stets auf eine Auswahl von Spracherzeugnissen angewiesen sein. Diese Auswahl muß dann repräsentativ sein, wenn über die bloße Aufzählung von Korpusbefunden oder die Angabe von Korpusbelegen zum Zwecke der Illustration hinaus Klassifizierungen und damit verbundene verallgemeinernde und erklärende Aussagen über den Objektbereich angestrebt sind.

Bevor wir uns einzelnen Fragen der Auswahl von sprachlichen Äußerungen für ein Textkorpus zuwenden, wollen wir zwei grundsätzliche Möglichkeiten des Charakters und damit auch der Zusammenstellung eines Korpus unterscheiden.

Entweder wird ein Korpus als eine Materialbasis verstanden, die vor allem dazu dienen soll, auf empirisch-deskriptive Weise u.a. auch Kriterien für Textdefinitionen und Textklassifikationen zu gewinnen. Die Auswahl sprachlicher Äußerungen für ein solches Korpus würde eine primär-induktive Korpustheorie voraussetzen. Im Rahmen einer solchen Theorie wären die notwendigsten Vorannahmen über den Objektbereich zu explizieren sowie Hypothesen über die Struktur dieses Objektbereichs zu formulieren. Um eine annähernde Repräsentativität der die Gesamtheit oder bestimmte Ausschnitte des Objektbereichs konstituierenden Strukturmuster zu gewährleisten, müßten zunächst Kategorien etabliert werden, die zum einen geeignet sind, den ganzen im Korpus zu repräsentierenden Sprachbereich zu erfassen, die zum anderen die Entscheidung darüber, welcher der Kategorien eine bestimmte sprachliche Äußerung zuzuordnen ist, weitgehend unabhängig vom subjektiven Ermessen des Auswählenden und damit auch eine Nachprüfbarkeit der jeweils getroffenen Entscheidung ermöglichen.¹²

Obwohl ein solches Korpus sehr umfangreich sein müßte, ist es vor allem wegen des notwendig vergleichsweise einfach zu handhabenden Kategorienschemas mit verhältnismäßig geringem Aufwand herstellbar. Es sei nicht verschwiegen, daß einem primär-induktiv begründetem Korpus ein vornehmlich heuristischer Wert zukommt, es sich neben seinen anderen Verwendungsmöglichkeiten aber auch dazu eignen könnte, als Ausgangs- oder Vergleichsmaterial für andere bzw. andersartige Korpora zu dienen, wie sie z.B. solche des im folgenden geschilderten Typus darstellen.

Oder aber das Korpus wird eher als eine Dokumentation verstanden, die den gesamten Objektbereich angemessen repräsentieren soll. Ein solches Korpus setzt eine primär-deduktive Korpustheorie voraus, in deren Rahmen die Prämissen über den Objektbereich zu formulieren sind. Dazu gehören u.a. auch die Definitionen der elementaren Einheiten und der Kate-

gorien ihrer Klassifikation. Ein solches Korpus würde das Ergebnis der aus den Prämissen gefolgerten Konklusion und damit gleichsam eine Illustration der die repräsentative Auswahl begründenden Korpusstheorie darstellen.

Gleichgültig welche Theorie der Erstellung eines Korpus zugrunde gelegt wird, so bestimmen auch im weiteren vorweg der wissenschaftstheoretische Standort des Korpuserstellers und sein Vorverständnis vom Objektbereich ebenso wie die beabsichtigte Verwendung des Korpus bzw. das Erkenntnisinteresse, welche Gesamtheit die vorzunehmende Auswahl repräsentieren soll.¹³ Zu entscheiden ist z.B., ob ein Sprachzustand oder eine Sprachentwicklung, geschriebene und/oder gesprochene Spracherzeugnisse, die Standardsprache und/oder diatopische und/oder diastratische Varietäten jene Gesamtheit bilden, aus der eine Auswahl getroffen werden soll.

Sind diese wesentlich durch das Erkenntnisinteresse geleiteten und zu begründenden Entscheidungen der Begrenzung des Objektbereichs gefallen, gilt es festzustellen, welche sprachlichen Einheiten als Elemente der so bestimmten Menge angenommen werden, und wie diese Einheiten definiert sind.

Gehen wir mit Hartmann (1971) 10 davon aus, daß "der Text, verstanden als grundsätzliche Möglichkeit des Vorkommens von Sprache in manifestierter Erscheinungsform, und folglich jeweils ein bestimmter Text als manifestierte Einzelercheinung funktionsfähiger Sprache, das originäre sprachliche Zeichen [bildet]", wäre im weiteren diese Einheit 'Text' im Rahmen einer Texttheorie bzw. Textgrammatik zu definieren.

Es ist hier nicht der Platz, die verschiedenen linguistischen, kommunikationstheoretischen und handlungstheoretischen Ansätze zu einer solchen Textdefinition vorzustellen und zu diskutieren.¹⁴ Wenn man die bisher erarbeiteten Ergebnisse sichtet, kommt man allerdings zu dem Schluß, daß die vorgelegten Definitionen entsprechend dem jeweiligen Entwicklungsstand der verschiedenen Texttheorien allenfalls als Entwürfe gelten dürfen. Für die zweifelsfreie Delimitation von Texten als elementaren Einheiten im Hinblick auf deren Zusammenstellung für ein Textkorpus eignen sie sich kaum, so daß für ein solches Vorhaben – bei aller zugestandenen Fragwürdigkeit dieses Vorgehens – allein bleibt, Texte aufgrund ethischer, d.h. textexterner Delimitationskriterien als Texte zu bestimmen.¹⁵

Um aus dem Gesamt von Texten, das Harweg (1968) 143 f. "Textkosmos" nennt, eine repräsentative Auswahl treffen zu können, ist es unerläßlich, die Menge aller Texte in Teilmengen zu gliedern bzw. das Gesamt aller Texte zu klassifizieren. Es wäre die definierte Menge aller individuellen Sprachrealisationen dann angemessen repräsentiert, wenn entsprechend

der Häufigkeit ihres festgestellten Vorkommens bzw. Gebrauchs die jeweiligen Teilmengen aufgrund bestimmter statistischer Kriterien durch entsprechend ausgewählte Textexemplare adäquat vertreten wären.

Zu den bekanntesten Textklassifikationssystemen zählen der semiotische Ansatz von Morris (1946) 125, das den kommunikativen Ansatz von Bühler (1969) und den semiotischen von Morris kombinierende Modell Kerns (1969) sowie die kommunikativ orientierten Vorschläge von Brinkmann (1971) 845 - 887 und Glinz (1971) 80 - 88.

Keines dieser sehr komplexen Klassifikationssysteme ist bei der Erstellung der hier dokumentierten Korpora zugrunde gelegt worden. Vielmehr handelt es sich hier meist um Klassifikationen nach literarischen Gattungen, nach Gebrauchsintentionen, Textstoffen und Publikationsformen, die einzeln oder in Kombination als Kriterien zur Charakterisierung der einzelnen Texte dienen. Nicht selten sind diese Kategorien erst nach Fertigstellung des Korpus ausgewählt worden, um eine Textauswahl nachträglich beschreiben, rechtfertigen oder auf ihre Relevanz hin überprüfen zu können.

3. Dokumentation von dreizehn maschinenlesbaren Textkorpora

Der Zweck unserer Dokumentation sollte nicht so sehr eine möglichst getreue Wiedergabe der mehr oder weniger ausführlichen Beschreibungen sein, die die im Anhang verzeichneten Publikationen über Umfang, Zusammensetzung und Zweck der einzelnen Textkorpora enthalten. Unsere Absicht war es vielmehr, eine vergleichende Übersicht über die bekanntesten derzeit maschinenlesbar zur Verfügung stehenden Textkorpora der deutschen Standardsprache der Gegenwart zu bieten.

Schwierigkeiten bei dieser Art der Darstellung bereiteten nicht so sehr die Angaben zur äußeren Gestalt der einzelnen Korpora. Sie waren entweder in den im Anhang zitierten Schriften enthalten oder durch Anfragen bei den Korpuserstellern zu erfahren.

Als problematisch erwies es sich allerdings, die zum Teil erheblich voneinander abweichenden Merkmale zur inhaltlichen Beschreibung der in die Korpora aufgenommenen Texte auf einen annähernd gleichen Nenner zu bringen. Im wesentlichen haben wir uns dabei an die von den Autoren selbst gewählten Kategorien gehalten.

Nicht aufgenommen haben wir in unserer Übersicht die Angabe, ob unter den Korpustexten (wie dies z.B. bei den beiden Aachener Korpora, beim Bonner und beim Saarbrücker Korpus der Fall ist) sich auch solche befinden, die Übersetzungen aus anderen Sprachen darstellen. Da Texte

solcher Art in allen genannten Korpora fehlen, haben wir in unserer Tabelle ebenfalls auf eine Kategorie "diatopische Varietät" verzichtet, im weiteren auch auf die nur für zwei Korpora ausgewiesene Kategorie "Textintention".

Zur Erläuterung der nachstehenden vergleichenden Übersicht muß noch erwähnt werden, daß die in Klammern gesetzten Daten angestrebte Zielvorstellungen angeben.

Wie schon an anderer Stelle gesagt wurde, hat die Zweckbestimmung einen entscheidenden Einfluß auf die Zusammensetzung des einzelnen Korpus. Wir möchten daher noch kurz darauf eingehen, zu welchem Zweck die einzelnen Korpora eingerichtet und bisher benutzt worden sind.

Das Lunder Zeitungskorpus ist in erster Linie für lexikalische Untersuchungen aufgebaut und bisher auch ausgewertet worden. Ähnlichen, d.h. im weiteren Sinne lexikalischen Analysen dienen auch die beiden Aachener Korpora, mit deren Hilfe auf der Basis statistischer Korrelationsanalysen maschinelle Verfahren der Feststellung semantischer Umgebungsfelder entwickelt worden sind.

Den Zwecken der Dokumentation, vergleichender Wortschatzuntersuchungen des öffentlichen Sprachgebrauchs in der Bundesrepublik und der DDR sowie der Erstellung eines lemmatisierten Wörterbuchs der ost-west-differenten Lexeme dient das Bonner Zeitungskorpus, das als Besonderheit gegenüber den anderen Korpora eine gleichmäßige Stufung der zeitlichen Herkunft seiner Texte aufweist.

Einen besonderen Status nimmt das Saarbrücker Korpus ein, das als Satzkorpus aufgebaut und mit einer Reihe syntaktischer Angaben per Hand codiert worden ist. Es liefert das Untersuchungs- und Testmaterial für das im Sonderforschungsbereich 100 der Universität des Saarlandes entwickelte maschinelle Syntaxanalyseverfahren.

Die weiteren Korpora sind vornehmlich als empirische Basis für grammatische Untersuchungen konzipiert und bisher ausgewertet worden. Die Anlehnung der Konzeption des LIMAS-Korpus an diejenige des Brown Corpus resultiert aus den damaligen Projektzielen der Forschungsgruppe LIMAS, ein maschinelles Übersetzungsverfahren Deutsch-Englisch-Deutsch zu entwickeln.¹⁶

Zusätzlichen didaktischen Zwecken dient die Auswertung der beiden Textkorpora Grunddeutsch von J. Alan Pfeffer. Primär für Unterrichtszwecke sind das Korpus der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts in München sowie das von Siliakus betreute Textkorpus DURF des Sprachlabors der Universität Adelaide zusammengestellt und bisher genutzt worden.

Es sei noch bemerkt, daß das Bonner Zeitungskorpus, das Freiburger Korpus, das LIMAS-Korpus, die beiden Textkorpora Grunddeutsch sowie das Korpus des Goethe-Instituts von vornherein auf eine repräsentative Auswahl von Texten aus dem jeweiligen Objektbereich ausgerichtet waren; ebenso das Lunder Zeitungskorpus, das sich durch eine hochgradige Aufnahme der Gesamttexte zweier bundesdeutscher Tageszeitungen auszeichnet. Das Mannheimer Korpus, in dem zunächst die Belletristik einen vergleichsweise großen Raum einnahm, ist durch bisher zu gering oder gar nicht vertretene Textarten mit dem Ziel ergänzt worden, eine Repräsentation der deutschen Standardschriftsprache der Gegenwart zu bieten.

Ein Problem eigener Art ist die Berücksichtigung von Fachtexten in einem Korpus. Das LIMAS-Kfz-Korpus besteht als einziges der in unsere Dokumentation aufgenommenen Korpora ausschließlich aus Texten eines Fachgebietes. Es ist hier nicht der Ort zu diskutieren, in welchem Sinn und in welchem Maß die Texte dieses Korpus fachsprachliche Texte sind, und auf welche Weise sie sich von Texten unterscheiden, die gemeinhin als standardsprachliche Texte bezeichnet werden. Wir meinen allerdings, daß die Hersteller der einzelnen Korpora eine Lösung des Problems der Berücksichtigung von fachsprachlichen Texten in einem Textkorpus insofern angeboten haben, als sie in einer Streuung, die von Korpus zu Korpus variiert, die verschiedensten Fach- bzw. Sach- bzw. Wissensgebiete repräsentiert haben.

Insgesamt gilt, daß die Frage der Repräsentativität einer Textauswahl bisher keine befriedigenden Antworten erfahren hat. Mit Sicherheit lassen sich ebenso viele Hypothesen über die Bedingungen der Möglichkeit der Repräsentativität einer Textauswahl formulieren, wie es unterschiedliche begründete wissenschaftstheoretische Standorte und Definitionen des Objektbereichs gibt. Die bisher erstellten Korpora bieten immerhin Konzeptionen und Materialien an, die bei einer Diskussion über dieses Thema nicht unberücksichtigt bleiben sollten.

Bleibt zum Schluß noch zu bemerken, daß die hier vorgelegte vergleichende Übersicht auch als Anregung zur Diskussion über ein verbindliches Muster für die Beschreibung von Korpora der in ihnen enthaltenen Texte dienen möge.

Tabellarische Übersicht über Umfang und Aufbau maschinenlesbarer Textkorpora des Deutschen (Stand: 1977)

		(1) Sprache	(2) nationale Varietät	(3) mediale Varietät	(4) diastratische Varietät	(5) Genre
1	Aachener Textkorpora: Prosatexte	Deutsch	BRD	geschrie- bene Texte	Standard- sprache	Romane, Er- zählungen, Sachbücher
2	Aachener Textkorpora: Studenten- lyrik	Deutsch	BRD	geschrie- bene Texte	Standard- und Litera- tursprache	Gedichte
3	Bonner Zeitungs- korpora	Deutsch	BRD DDR	geschrie- bene Texte	Standard- sprache	Zeitungs- texte
4	Textkorpora Grunddeutsch Schreib- sprache	Deutsch	BRD DDR Schweiz Österreich	geschrie- bene Texte	Standard- sprache	Sachbücher, Berichte, Ro- mane, Er- zählungen
5	Textkorpora Grunddeutsch Sprech- sprache	Deutsch	BRD DDR Schweiz Österreich	gespro- chene Texte	Standard- sprache	Erzählungen, Berichte, Un- haltungen, Interviews
6	Freiburger Korpora	Deutsch	BRD	gespro- chene Texte	Standard- sprache	Dialoge, Monologe
7	Adelaide- Korpora DURF	Deutsch	BRD	geschrie- bene Texte	Fach- sprachen	Lehrbücher, Berichte, Sachbe- schreibungen, Abhand- lungen
8	LIMAS- Korpora	Deutsch	BRD	geschrie- bene Texte	Standard- sprache	Sachbücher, Erzählungen, Berichte, Romane

(6) Publika- tionsform	(7) Text- stoff	(8) zeitliche Herkunft der Texte	(9) Anzahl der Texte	(10) Anzahl der Wortstellen pro Text	(11) Anzahl der Wortstellen des Korpus
Buch	fiktional, nicht- fiktional	1966-1970	42	32.000 - 295.000	4 Mio
Buch	fiktional	1822-1962	4.000	12-480	520.000
Zeitung	nicht- fiktional	1949 1964 1954 1969 1959 1974	18.000	3 Zeilen- 999 Zeilen	4 Mio
Buch	fiktional, nicht-fik- tional (23 Wissens- gebiete)	1964-1965	500	1.000	500.000
Dialog	nicht- fiktional (25 vorge- gebene Ge- sprächsthe- men)	1960-1961	400	1.500	595.000
Vortrag, Gespräch	nicht- fiktional (9 Sach- gebiete)	1967-1974	222	175- 16.360	600.000
Buch	nicht- fiktional	1950-1965	70	10.000	650.000
Buch, Zeit- schrift, Zeitung	fiktional, nicht- fiktional (34 Sachge- biete)	1970	500	2.000	1 Mio

Tabellarische Übersicht (Fortsetzung):

		(1) Sprache	(2) nationale Varietät	(3) mediale Varietät	(4) diastratische Varietät	(5) Genre
9	LIMAS Kfz-Korpus	Deutsch	BRD	geschrie- bene Texte	Fach- sprache	Lehrbücher, Betriebsan- weisungen, Berichte
10	Lunder Zeitungs- Korpus	Deutsch	BRD	geschrie- bene Texte	Standard- sprache	Zeitungs- texte
11	Mannheimer Korpus	Deutsch	BRD DDR Schweiz Österreich	geschrie- bene Texte	Standard- und Litera- tursprache	Romane, Er- zählungen, Berichte
12	Saarbrücker Korpus	Deutsch	BRD	geschrie- bene Texte	Standard- und Fach- sprache	Fachbücher, Zeitungs- texte
13	Textkorpus des Goethe- Instituts	Deutsch	BRD	geschrie- bene Texte	Fach- sprachen	Lehrbücher, Berichte, Sachbe- schreibungen, Abhand- lungen

(6) Publika- tionsform	(7) Text- stoff	(8) zeitliche Herkunft der Texte	(9) Anzahl der Texte	(10) Anzahl der Wortstellen pro Text	(11) Anzahl der Wortstellen des Korpus
Buch, Zeit- schrift	nicht- fiktional (Sachgebiet Kfz-Wesen)	1956-1974	90 (200)	600 - 77.500	750.000 (1 Mio)
Zeitung	nicht- fiktional (5 Sach- gebiete)	1966-1967	6.595	?	3 Mio
Buch, Zeit- schrift, Zeitung	fiktional, nicht- fiktional (10 Sach- gebiete)	1946-1967	32	7.800 - 144.000	2,2Mio
Buch, Zeitung	nicht- fiktional	1955-1957 1961-1963	11.000 Sätze aus 45 Buch- u. 1.340 Zei- tungstexten	?	200.000
Buch	nicht- fiktional (34 Fach- gebiete)	1950-1967	102 (pro Fach 3)	?	250.000

4. Verzeichnis der Publikationen über die dokumentierten Textkorpora und ihre Verwendung

1. Aachener Textkorpora: Prosatexte
 - 1.1. Dannhauer, H.M./Wickmann, D. (1972): Die Segmentierung fortlaufender Prosatexte zur Bestimmung semantischer Umgebungsfelder. In: Zs. für Literaturwissenschaft und Linguistik 2, Heft 8, 44 - 49.
2. Aachener Textkorpora: Studentenlyrik
 - 2.1. Rieger, B. (1970): *Poetae Studiosi*. Frankfurt/M.
 - 2.2. Rieger, B. (1971): Wort- und Motivkreis als Konstituenten lyrischer Umgebungsfelder. In: Lili 1, Heft 4, 23 - 41.
 - 2.3. Rieger, B. (1972): Über die Erstellung eines synoptischen Affinitätslexikons (SAL). In: H. Schanze (Hg.): *Literatur und Datenverarbeitung*. Tübingen. 123 - 130.
 - 2.4. Rieger, B. (1972): Literarische Massenphänomene und mengenorientierte Textanalyse. In: Motte-Haber (Hg.): *Das Trivale in Musik, Literatur und bildender Kunst*, Frankfurt/M. 42 - 62.
 - 2.5. Rieger, B. (1972): Warum mengenorientierte Textwissenschaft? Zur Begründung der Statistik als Methode. In: LiLi 8, 11 - 28.
 - 2.6. Rieger, B. (1974): Eine tolerante Lexikonstruktur. Zur Abbildung natürlich-sprachlicher Bedeutung auf unscharfe Mengen in Toleranzräumen. In: LiLi 16, 31 - 47.
 - 2.7. Rieger, B. (1977): Vagheit als Problem der linguistischen Semantik. Zur Verbindung von Methoden der Statistik mit der Theorie der unscharfen Mengen bei der Analyse von Bedeutung. In: Bald, W.D. (Hg.): *Semantik und Pragmatik. Akten des 11. Linguistischen Kolloquiums*. Tübingen. 91 - 101.
 - 2.8. Rieger, B. (1977): Theorie der unscharfen Mengen und empirische Textanalyse. In: Klein, W. (Hg.): *Methoden der Textanalyse. Medium Literatur 3*. Heidelberg. 84 - 99.
3. Bonner Zeitungskorpus
 - 3.1. Hellmann, M.W. (1968): Zur Dokumentation und maschinellen Bearbeitung von Zeitungstexten in der Außenstelle Bonn. In: *Forschungsberichte des Instituts für deutsche Sprache 2*. Mannheim. 39 - 125.
 - 3.2. Hellmann, W.M. (1973): Bericht über die Arbeit der Forschungsstelle Bonn des IdS. In: M.W. Hellmann (Hg.): *Zum öffentlichen Sprachgebrauch in der Bundesrepublik Deutschland und in der DDR (= Sprache der Gegenwart 18)*. Düsseldorf. 15 - 34.

- 3.3. Hellmann, M.W. (1976): Möglichkeiten und Probleme bei vergleichenden Wortschatzuntersuchungen zum öffentlichen Sprachgebrauch in der Bundesrepublik und der DDR. In: Probleme der Lexikologie und Lexikographie (= Sprache der Gegenwart 39). Düsseldorf. 242 - 274.
- 3.4. Schaefer, B. (1978): Das Bonner Zeitungskorpus. Eine maschinelle Dokumentation von Tageszeitungen der BRD und der DDR. Mimeo. Bonn.
4. Textkorpus Grunddeutsch (Schriftsprache)
 - 4.1. Pfeffer, J.A. (1962): Grunddeutsch – Werden und Wesen. In: The German Quarterly 31.
 - 4.2. Pfeffer, J.A. (1969): Grunddeutsch. In: Zeitschrift für deutsche Sprache 25.
 - 4.3. Pfeffer, J.A. (1972): Das Institut für Grunddeutsch. In: Jahrbuch für Internationale Germanistik 3.
 - 4.4. Pfeffer, J.A. (1975): Grunddeutsch. Erarbeitung und Wertung dreier deutscher Korpora (= Forschungsberichte des Instituts für deutsche Sprache 27). Tübingen.
5. Textkorpus Grunddeutsch (Sprechsprache)

Die bereits unter "Textkorpus Grunddeutsch (Schriftsprache)" genannte Literatur wird hier nicht noch einmal aufgeführt.

 - 5.1. Pfeffer, J.A. (1963): Evolution of the Basic (Spoken) German Word List. Level I. Pittsburgh.
 - 5.2. Pfeffer, J.A. (1964): Basic (Spoken) German Word List. Level I. Englewood Cliffs, New Jersey.
 - 5.3. Pfeffer, J.A. (1965): Index of English Equivalents für the Basic (Spoken) German Word List. Level I. Englewood Cliffs, New Jersey.
 - 5.4. Pfeffer, J.A. (1968): Basic (Spoken) German Idiom List. Level I. Englewood Cliffs, New Jersey.
 - 5.5. Pfeffer, J.A. (1970): Basic (Spoken) German Word List. Level II. Pittsburgh.
 - 5.6. Pfeffer, J.A. (1970): Basic (Spoken) German Dictionary for Everyday Usage. Englewood Cliffs, New Jersey.
6. Freiburger Korpus gesprochener Texte des Deutschen
 - 6.1. Steger, H. (1967): Gesprochene Sprache. Zu ihrer Typik und Ter-

- minologie. In: Satz und Wort im heutigen Deutsch (= Sprache der Gegenwart 1). Düsseldorf. 259 - 291.
- 6.2. Steger, H. (1970): Über Dokumentation und Analyse gesprochener Sprache. In: Zielsprache Deutsch, Heft 1, 13 - 21; Heft 2, 51 - 63.
- 6.3. Steger, H. u.a. (Hg.) (1971 - 75): Texte gesprochener deutscher Standardsprache 1 - 3. München, Düsseldorf.
- 6.4. Engel, U./Vogel, I. (Hg.) (1973): Gesprochene Sprache (= Forschungsberichte des Instituts für deutsche Sprache 7). Mannheim.
7. Adelaide-Korpus DURF
- 7.1. Siliakus, H. (1970): Some Reflexions on the Lack of Accuracy of Word Frequency Lists (with K. Morris). In: ITL 9, 11 - 19.
- 7.2. Siliakus, H. (1976): Erwägungen bei der Zusammenstellung eines Grundwortschatzes — ein Forschungsbericht. In: Deutsch als Fremdsprache 6, 367 - 372.
8. LIMAS-Korpus
- 8.1. Hoppe, A./Schader, B. (1973): Das LIMAS-Korpus. Ein Korpus geschriebener Texte der deutschen Gegenwartssprache. Bonn (vervielf. Manuskript).
- 8.2. Glas, R. (1974): Ein Textkorpus für deutsche Gegenwartssprache. In: IRAL-Sonderband GAL'73. Heidelberg. 112 - 115.
- 8.3. Glas, R. (1975): Das LIMAS-Korpus, ein Textkorpus für die deutsche Gegenwartssprache. In: Linguistische Berichte, Heft 40, 63 - 66.
9. LIMAS-Kfz-Korpus
- 9.1. Schader, B. (1976): Das LIMAS-Kfz-Korpus. Bonn (vervielf. Manuskript).
10. Lunder Zeitungskorpus
- 10.1. Rosengren, I. (1968): Ein Frequenzwörterbuch der modernen Zeitungssprache — wie und wozu? In: Beiträge zur Linguistik und Informationsverarbeitung. Heft 14, 7 - 21.
- 10.2. Rosengren, I. (1972): Ein Frequenzwörterbuch der deutschen Zeitungssprache. DIE WELT — Süddeutsche Zeitung. Bd. 1. Lund.
- 10.3. Rosengren, I. (1978): Ein Frequenzwörterbuch der deutschen Zeitungssprache. DIE WELT — Süddeutsche Zeitung. Bd. 2. Lund (i. Vorb.).

11. Mannheimer Korpus

- 11.1. Zint, I. (1968): Maschinelle Sprachbearbeitung des Instituts für deutsche Sprache in Mannheim. In: Forschungsberichte des Instituts für deutsche Sprache 2. Mannheim. 9 - 38.
- 11.2. Billmeier, G. (1968): Über die Signifikanz von Auswahltexten. In: Forschungsberichte des Instituts für deutsche Sprache 2. Mannheim. 126 - 171.
- 11.3. Hellmann, M.W. (1969): Über Corpusgewinnung und Dokumentation im Mannheimer Institut für deutsche Sprache. In: Forschungsberichte des Instituts für deutsche Sprache 3. Mannheim. 25 - 54.
- 11.4. Engel, U. (1969): Das Mannheimer Corpus. In: Forschungsberichte des Instituts für deutsche Sprache 3. Mannheim. 75 - 84.
- 11.5. Bayer, K./Kurbel, K./Epp, B. (1974): Maschinelle Textverarbeitung im Rechenzentrum des IdS (= Forschungsberichte des Instituts für deutsche Sprache 14). Mannheim.

12. Saarbrücker Korpus

- 12.1. Eggers, H. (1962): Zur Syntax der deutschen Sprache der Gegenwart. In: Studium Generale 15. 49 - 59.
- 12.2. Eggers, H. u.a. (1969): Elektronische Syntaxanalyse der deutschen Gegenwartssprache. Tübingen (S. 3 - 5: Das Sprachmaterial).
- 12.3. Zimmermann, H. u.a. (1974): Auswertung des Saarbrücker Korpus. In: Arbeitsbericht 1972 - 1974 des Sonderforschungsbereichs Elektronische Sprachforschung. Projektbereich A: Germanistik. Universität des Saarlandes. Saarbrücken. 359 - 374.

13. Textkorpus des Goethe-Instituts

- 13.1. Erk, H. (1972): Zur Lexik wissenschaftlicher Fachtexte. Verben - Frequenz und Verwendungsweise (= Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts, Band 4). München.
- 13.2. Erk, H. (1975): Zur Lexik wissenschaftlicher Fachtexte. Substantive - Frequenz und Verwendungsweise (= Schriften der Arbeitsstelle für wissenschaftliche Didaktik des Goethe-Instituts, Band 5). München.

5. Verzeichnis der Anschriften der für die Herstellung und für die Vergabe der einzelnen Textkorpora zuständigen Stellen

1. Aachener Textkopus: Prosatexte

Dr. H.M. Dannhauer
Technische Hochschule Aachen
Germanistisches Institut
Templergraben 55
5100 Aachen

2. Aachener Textkopus: Studentenlyrik

Dr. B. Rieger
Technische Hochschule Aachen
Germanistisches Institut
Templergraben 55
5100 Aachen

3. Bonner Zeitungskopus

Dr. M.W. Hellmann
Institut für deutsche Sprache
Forschungsstelle Bonn
Kaiserstr. 46
5300 Bonn

4. Textkopus Grunddeutsch (Schriftsprache)

Prof. Dr. J. Alan Pfeffer
Department of German Studies
Institute for Basic German
Stanford University
Stanford, California 94305

5. Textkopus Grunddeutsch (Sprechsprache)

siehe unter Textkopus Grunddeutsch (Schriftsprache)

6. Freiburger Kopus gesprochener Texte der Deutschen

Professor Dr. H. Steger
Institut für deutsche Sprache
Forschungsstelle Freiburg
Belfortstr. 14
7800 Freiburg

7. Adelaide-Korpus DURF
Henk J. Siliakus, M.A.
Director of the Language Laboratory
University of Adelaide
Adelaide
Australia, 5000
8. LIMAS-Korpus
Prof. Dr. W. Lenders
Institut für Kommunikationsforschung
und Phonetik der Universität Bonn
Poppelsdorfer Allee 49
5300 Bonn 1
9. LIMAS-Kfz-Korpus
siehe unter LIMAS-Korpus
10. Lunder Zeitungskorpus
Prof. Dr. I. Rosengren
Germanistisches Institut der
Universität Lund
Helgonabacken 14
S-22362 Lund
11. Mannheimer Korpus
P. Wolfangel, M.A.
Institut für deutsche Sprache
Friedrich-Karl-Str. 12
6800 Mannheim
12. Saarbrücker Korpus
Prof. Dr. H. Eggers
Germanistisches Institut der
Universität des Saarlandes
6600 Saarbrücken
13. Textkorpus des Goethe-Instituts
Dr. H. Erk
Goethe-Institut
Arbeitsstelle für wissenschaftliche Didaktik
Ohmstr. 8
8000 München 40

Anmerkungen

- 1 Vgl. Mentrup (1976).
- 2 Eine Dokumentation auch anderssprachiger Textkorpora ist in Bergenholtz/Schaeder (1978) enthalten. Obwohl sie annähernd 4 Mio. Wortstellen umfaßt, ist die sogenannte LIMAS-Textsammlung nicht in diese Dokumentation aufgenommen worden, weil sie lediglich ein Konglomerat von Texten des Aachener Textkorpus: Prosatexte, des LIMAS-Korpus sowie einiger Zeitungen darstellt. Eine Beschreibung der LIMAS-Textsammlung ist in Schaeder (1977) enthalten.
- 3 Erst diese spezifische Differenz ermöglicht die in den meisten Definitionen vernachlässigte Unterscheidung von Textkorpora und beliebigen Textsammlungen, wie sie z.B. auch Märchenbücher, Festschriften und Sammelbände sonstiger Art darstellen.
- 4 Vgl. hierzu Petöfi/Rieser (1974); van de Velde (1974), (1974a).
- 5 Bierwisch (1966) 5.
- 6 Chomsky (1969) 14.
- 7 Chomsky (1969) 15.
- 8 Bierwisch (1966) 9.
- 9 Bierwisch (1966) 9.
- 10 Leech (1971) 8.
- 11 Vgl. hierzu van de Velde (1974a) 32 - 36.
- 12 Vgl. Glas (1974) 64.
- 13 Vgl. Schank (1974) 20 f.
- 14 Als Beispiel für eine linguistische Textdefinition vgl. Harweg (1968), als Beispiel für eine handlungstheoretische vgl. Kummer (1972), als Beispiel für eine kommunikationstheoretische vgl. Schmidt (1973). Eine vergleichende Darstellung einer Auswahl von Textmodellen bieten Gülich/Raible (1977).
- 15 Zur Terminologie, zur Kritik etischer sowie zum Versuch emischer Textdefinition vgl. Harweg (1968) 139 - 178.
- 16 Zum Brown Corpus vgl. Kučera/Francis (1967).
 Während im Brown Corpus die ausgewählten Texte nach alternativen Merkmalen, wie z.B. "informative prose/imaginative prose" oder "press/books/periodicals" klassifiziert sind, wurde für das LIMAS-Korpus durchgehend eine Klassifikation nach Sachgebieten gewählt. Die Sachgebietsklassifikation ist an derjenigen der Deutschen Bibliographie orientiert. Die Gewichtung der Sachgebiete richtet sich nach der Anzahl der 1971 in der Deutschen Bibliographie pro Sachgebiet verzeichneten Publikationen. Siehe Glas (1975).
 Ohne daß dies den Herstellern des LIMAS-Korpus bekannt gewesen ist, war früher bereits J. Alan Pfeffer ähnlich verfahren, als er für die Auswahl der Texte seines Korpus Grunddeutsch (Schriftsprache) eben die Sachgebietsklassifikation der Deutschen Bibliographie zugrunde gelegt und die Gewichtung der Texte pro Sachgebiete in derselben Weise vorgenommen hatte, wie sie später auch für die Zusammenstellung des LIMAS-Korpus erfolgte.

Literatur

In der Arbeit zitierte und hier nicht aufgeführte Titel sind im "Verzeichnis der Publikationen über die dokumentierten Textkorpora und ihre Verwendung" (Kapitel 4) zu finden.

- Andresen, H. (1974): Der Erklärungsgehalt linguistischer Theorien (= Linguistische Reihe 18). München.
- Bausch, K.-H. (1975): Zur Problematik der empirischen Basis in der Linguistik. In: Zeitschrift für germanistische Linguistik, Heft 2, 123 - 148.
- Belke, H. (1973): Literarische Gebrauchsformen (= Grundstudium Literaturwissenschaft 9). Düsseldorf.
- Bergenholtz, H. / Schaefer, B. (1978) (Hg.): Text-Corpora. Materialien für eine empirische Sprach- und Literaturwissenschaft. Kronberg/Ts.
- Bierwisch, M. (1966): Grammatik des deutschen Verbs (= Studia Grammatica 2). 3. Aufl. Berlin.
- Brinkmann, H. (1971): Die deutsche Sprache. Gestalt und Leistung. 2, neubearb. und erw. Aufl. Düsseldorf.
- Bühler, K. (1969): Die Axiomatik der Sprachwissenschaften. Frankfurt/M.
- Bungarten, Th. (1973): Ein Korpus für Untersuchungen über die geschriebene deutsche Gegenwartssprache. In: Bungarten: Sprache und Sprachanalyse des Deutschen. Vier Beiträge zur Methode und Theorie (= Europäische Hochschulschriften. Reihe 1: Deutsche Literatur und Germanistik 76). Bern, Frankfurt/M., 73 - 97.
- (1976): Bemerkungen zur Rolle des Korpus und einige Kriterien zur Textklassifikation. In: Linguistica Biblica, Heft 38, 81 - 95.
- Chomsky, N. (1969): Aspekte der Syntaxtheorie. Aus dem Amerikanischen übersetzt und herausgegeben von einem Kollektiv unter der Leitung von E. Lang. Frankfurt/M.
- Filipović, R. (1974): The Use of a Corpus in Contrastive Studies. In: Raabe, H. (Hg.): Trends in kontrastiver Linguistik I (= Forschungsberichte des Instituts für deutsche Sprache 16). Mannheim. 51 - 66.
- Glinz, H. (1971): Soziologisches im Kernbereich der Linguistik. In: Sprache und Gesellschaft (= Sprache der Gegenwart 13). Düsseldorf. 80 - 88.
- Gülich, E. / Raible, W. (1977): Linguistische Textmodelle (= UTB 130). München.
- Hartmann, P. (1971): Texte als linguistisches Objekt. In: W.D. Stempel (Hg.): Beiträge zur Textlinguistik. München. 9 - 29.
- Harweg, R. (1968): Pronomina und Textkonstitution (= Beihefte zu Poetica 2). München.
- Kern, P. (1969): Bemerkungen zum Problem der Textklassifikation. In: Forschungsberichte des Instituts für deutsche Sprache 3. Mannheim. 3 - 23.
- Kučera, H. / Francis, W.N. (1967): Computational Analysis of Present-Day American English. Providence, Rhode Island, Brown University.

- Kummer, W. (1972): Aspects of a theory of argumentation. In: E. Gülich/
W. Raible (Hg.): Textsorten. Differenzierungskriterien aus linguistischer
Sicht (= Athenäum-Skripten Linguistik 5). Frankfurt/M. 25 - 49. (2. Aufl.
1975).
- Leech, G.N. (1971): Towards a Semantic Description of English. 2. Aufl. London.
- Leonard, R. (1973): The Computer Archive of Modern English Texts (The Lan-
caster Corpus). Preprint No. 34, International Conference on Computational
Linguistics, Pisa.
- Levelt, W.J.M. (1972): Some psychological aspects of linguistic data. In:
Linguistische Berichte, Heft 17, 18 - 30.
- Mentrup, W. (1976): Projekt eines großen interdisziplinären Wörterbuchs der deutschen
Sprache. In: Deutsche Sprache, Heft 1, 93 - 96; Heft 2, 188 - 190; Heft 4,
370 - 379.
- Morris, Ch. W. (1946): Signs, Language, and Behavior. New York.
- Nikitopoulos, P. (1974): Vorgriffe auf eine Thematisierung der Repräsentativität
eines Corpus. In: Deutsche Sprache, Heft 1, 32 - 42.
- Petöfi, J.S./Rieser, H. (1974): Wissenschaftstheoretische Argumente für eine um-
fassende grammatische Theorie und eine logisch-semantische Beschreibungs-
sprache. In: M. Rüttenauer (Hg.): Textlinguistik und Pragmatik (= Papiere
zur Textlinguistik 3). Hamburg. 1 - 46.
- Schader, B. (1976): Maschinenlesbare Textkorpora des Deutschen und des Eng-
lischen. In: Deutsche Sprache, Heft 4, 356 - 370.
- (1977): Die LIMAS-Textsammlung. Bonn (vervielf. Manuskript).
- Schank, G. (1974): Zur Korpusfrage in der Linguistik. In: Deutsche Sprache, Heft 4,
16 - 26.
- Schmidt, S.J. (1973): Texttheorie. Probleme einer Linguistik der sprachlichen
Kommunikation. München.
- Toulmin, St. (1958): The Uses of Argument. London.
- Velde, R.G. van de (1974): Linguistische Theoriesprache und Textlinguistik. In:
M. Rüttenauer (Hg.): Textlinguistik und Pragmatik (= Papiere zur Textlingui-
stik 3). Hamburg. 47 - 61.
- (1974a): Zur Theorie der linguistischen Forschung. München.
- Wunderlich, D. (1976)(Hg.): Wissenschaftstheorie der Linguistik. Kronberg/Ts.